Normative modeling of auditory memory for natural sounds

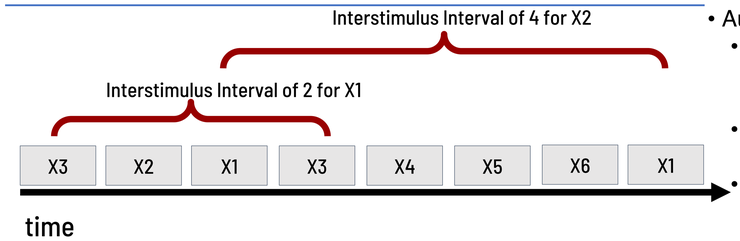
Bryan J. Medina[1,2,3] & Josh H. McDermott[1,2,3,4]

1. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
2. McGovern Institute for Brain Research, Massachusetts Institute of Technology
3. Center for Brains, Minds, and Machines, Massachusetts Institute of Technology
4. Program in Speech and Hearing Biosciences and Technology, Harvard University

Introduction

- Auditory memory is integral to everyday life, but underlying representations remain poorly understood
- **Goal: develop a normative account of memory of actual sensory signals**
- Further understand:
  - why some sounds are forgettable and some are memorable,
  - why false memories occur,
  - and why memory is the way it is.

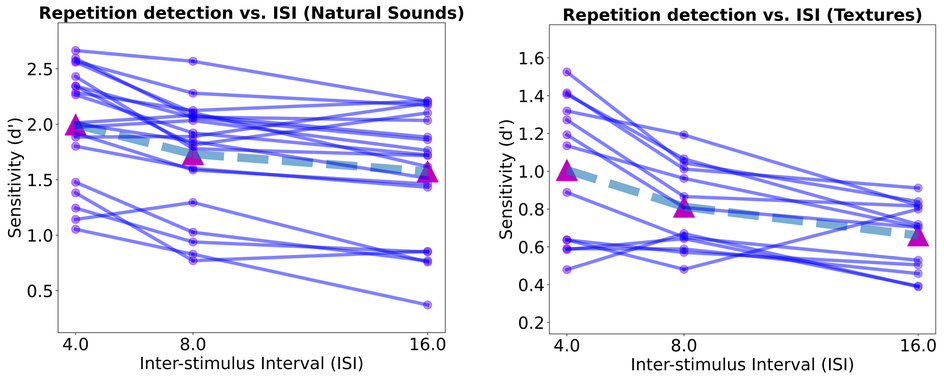
Psychophysical Methods



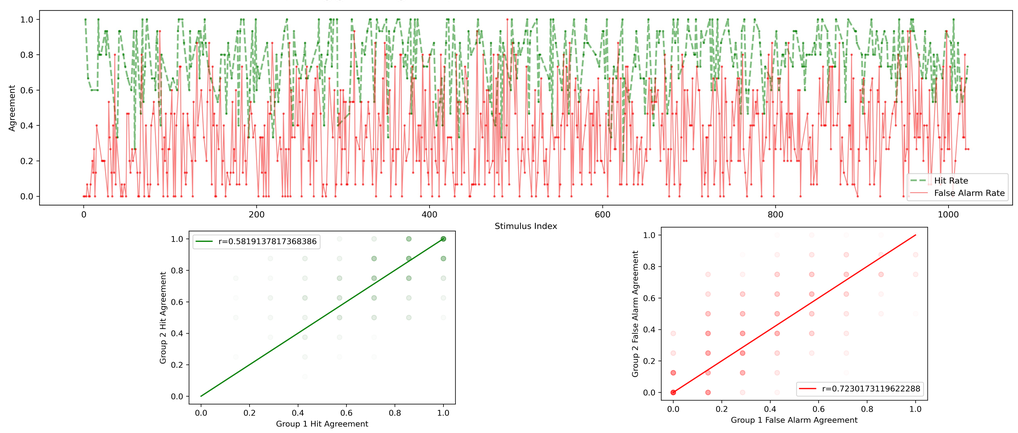
- Auditory memory experiment:
  - 2 second clips of natural auditory scenes, serially presented
  - Some clips repeat within experiment, some don't
  - After each presentation, participant asked if clip was heard before in the experiment
- Main experimental parameter:
  - Interstimulus Interval (ISI); number of intervening stimuli between a first presentation and its repetition

Psychophysical Results

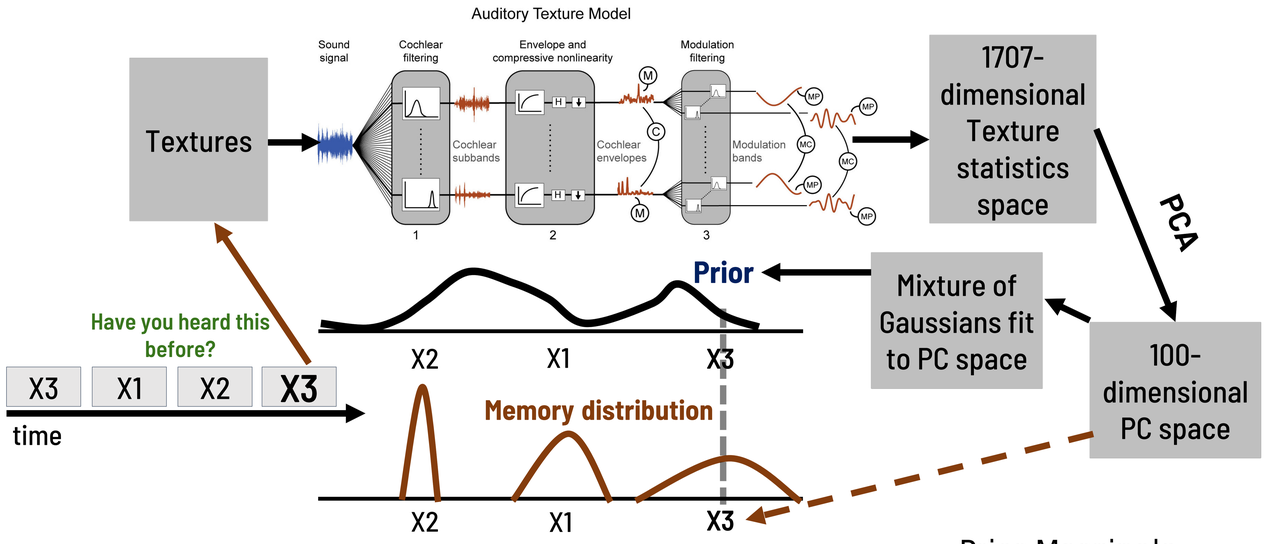
- In an experiment where clips can only repeat at most once, performance decreases as the interstimulus interval increases.
- Repeated experiment with auditory textures.
- Performance trends consistent across both experiments



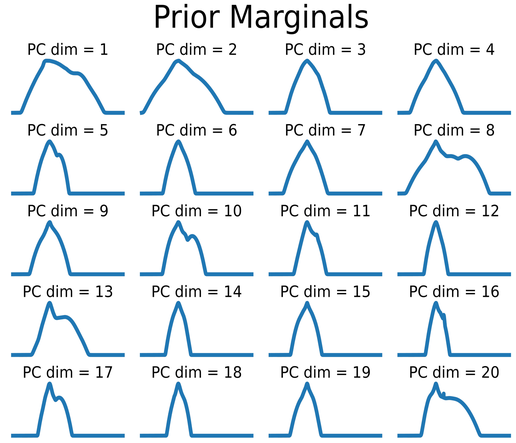
- When multiple participants viewed same stimulus sequence, there is agreement on hits and false alarms, suggesting responses are not random



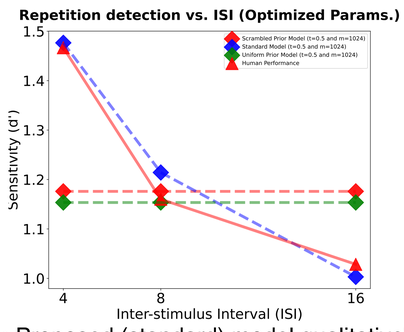
Ideal Observer Methods

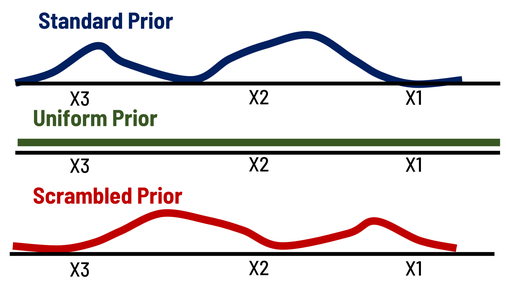


- Ideal observer model encodes experiences as memory traces whose noise grows with time
- Observer decides whether a subsequent stimulus is more likely to have come from the distribution of stimuli implied by these memory traces or from the prior distribution over all textures.
- Memory traces assumed to be based on texture statistics of heard stimulus
- ~143,000 textures from natural auditory scenes (obtained from the AudioSet dataset).
- Ideal observer is assumed to have knowledge of how noise in a trace grows with time. The memory traces are thus associated with a distribution of experiences, modeled as a mixture of Gaussians in the space of texture statistics.
- The prior is modeled as a mixture of Gaussians fit to the principal components of the statistics of a large set of natural textures.



Ideal Observer Results





- To determine necessity of prior for replicating human judgments, proposed model is compared to models with alternative priors:
  - Uniform prior over texture statistics
  - "Scrambled" priors with shuffled means and covariances
- Proposed (standard) model qualitatively mirrors human trends on the same task, while alternative models do not.
- Suggests that humans have internalized a prior over stimulus statistics in memory and use that to make memory judgments

Future work

- Observe and model the effects that number of stimulus repetitions has on humans recognition performance
- Use model to predict human judgments on a trial by trial basis
- Use recent advances in deep neural networks to obtain a representation for all natural sounds, and not just auditory textures